Iraqi Journal of Statistical Science (20) 2011
The Fourth Scientific Conference of the College of Computer Science & Mathematics
pp [202-213]

Modified Conjugate Gradient Algorithm with Proposed Conjugancy Coefficient

Dr.Abdul-Ghafoor J. Salem*. Dr.Ban A. Metras**
Nazar Kh. Hussian***

Abstract

In this paper, we present modified conjugancy coefficient for the conjugate gradient method based on the (Liu and Storey) method to solve non-linear programming problems. We proved the sufficient descent and the global convergence properties for the proposed algorithm for three cases and we get very good numerical results especially for the large scale optimization problem.

تحسين خوارزمية التدرج المترافق مع معامل ترافق مقترح

الملخص

تم في البحث اشتـقاق معامل ترافق محسن لطريقة المتـجهات المترافقة أساسه صيغة Liu وStorey لحل مسائل البرمجة غير الخطية وتم إثبات خاصية الانـحدار الكافي (sufficient descent) وخاصية التقارب الشامل للخوارزمية المقترحة بثلاث حالات، كما تم الحصول على نتائج عددية جيدة جدا وخاصة لمسائل الأمثلية ذات القياس العالي.

Keyword: conjugate gradient, conjugancy coefficient, nonlinear programming, unconstrained optimization .

1. Introduction

In unconstrained optimization, we minimize an objective function which depends on real variables with no restrictions on the values of these variables. The unconstrained optimization problem is:

$$Min \quad f(x) : x \in R^n, \qquad (1)$$

where $f : R^n \to R$ is a continuously differentiable function, bounded from below. A nonlinear conjugate gradient method generates a sequence $\{x_k\}, k$ is integer number, $k \geq 0$. Starting from an

*Prof. \ College of Computers Sciences and Math.\ University of Mosul
**Assistant Prof. \ College of Computers Sciences and Math.\ University of Mosul.
***Lecturre \ College of Computers Sciences and Math.\ University of Tikreet.\

initial point x_0, the value of x_k is calculated by the following equation:

$$x_{k+1} = x_k + \lambda_k d_k , \tag{2}$$

where the positive step size $\lambda_k > 0$ is obtained by a line search, and the directions d_k are generated as:

$$d_{k+1} = -g_{k+1} + \beta_k d_k , \tag{3}$$

where $d_0 = -g_0$, the value of β_k is determined according to the algorithm of Conjugate Gradient (CG), and its known as a conjugate gradient parameter, $s_k = x_{k+1} - x_k$ and $g_k = \nabla f(x_k) = f'(x_k)$, consider $\|.\|$ is the Euclidean norm and $y_k = g_{k+1} - g_k$. The termination conditions for the conjugate gradient line search are often based on some version of the Wolfe conditions. The standard Wolfe conditions:

$$f(x_k + \lambda_k d_k) - f(x_k) \le \rho \lambda_k g_k^T d_k , \tag{4}$$

$$g(x_k + \lambda_k d_k)^T d_k \ge \sigma g_k^T d_k , \tag{5}$$

where d_k is a descent search direction and $0 < \rho \le \sigma < 1$, where β_k is defined by one of the following formulas:

$$\beta_k^{(HS)} = \frac{y_k^T g_{k+1}}{y_k^T d_k} (\textit{Hestenese and Stiefel}[6]) \tag{6}$$

$$\beta_k^{(FR)} = \frac{g_{k+1}^T g_{k+1}}{g_k^T g_k} (\textit{Fletcher and Reeves}[4]) \tag{7}$$

$$\beta_k^{(PRP)} = \frac{y_k^T g_{k+1}}{g_k^T g_k} (\textit{Polak} - \textit{Ribiere } [8] \textit{and Polyak } [9]) \tag{8}$$

$$\beta_k^{(CD)} = -\frac{g_{k+1}^T g_{k+1}}{g_k^T d_k} (\textit{Conjugate descent } [5]) \tag{9}$$

$$\beta_k^{(LS)} = -\frac{y_k^T g_{k+1}}{g_k^T d_k} (\textit{Liu and Stoery } [10]) \tag{10}$$

$$\beta_k^{(DY)} = \frac{g_{k+1}^T g_{k+1}}{y_k^T s_k} (\textit{Dai and Yuan } [2]) \tag{11}$$

May not often have better computational performances. In order to exploit the attractive feature of each set, the so called new conjugate gradient method has been proposed as:

2. The Proposed Conjugate Gradient Algorithm:

The proposed algorithm generates the iterate $x_0, x_1, x_3, \ldots, x_n$ compute by the equation (2), the step size $\lambda_k > 0$ is determined according to the Wolfe conditions (4) and (5), and the direction d_k are generated by the equation (3):

The Liu and Storey Conjugate Gradient algorithm is one of the best methods to solve the large scale non-linear optimization

problems, we proposed the new algorithm by using the Liu and Storey formula as:
Since

$$\beta_k^{(LS)} = -\frac{g_{k+1}^T y_k}{g_k^T d_k} \tag{12}$$

$$= -\frac{g_{k+1}^T (g_{k+1} - g_k)}{g_k^T d_k}$$

$$= -\frac{g_{k+1}^T g_{k+1} - g_{k+1}^T g_k}{g_k^T d_k}$$

$$= -\frac{\|g_{k+1}\|^2}{g_k^T d_k} + \frac{g_{k+1}^T g_k}{g_k^T d_k}$$

By using the relation ($u^T v = \|u\| \ \|v\| \ \cos(\theta)$; where θ is the angle between the vectors u and v) we get:

$$= -\frac{\|g_{k+1}\|^2}{g_k^T d_k} + \frac{\|g_{k+1}\|\|g_k\|\cos(\theta_1)}{\|g_k\|\|d_k\|\cos(\theta_2)}$$

where θ_1, θ_2 are the angles between g_{k+1}, g_k and d_k, g_k respectively.

$$= \beta_k^{(CD)} + \frac{\|g_{k+1}\|\|g_k\|}{\|g_k\|\|d_k\|}\omega, \tag{13}$$

where:

$$\omega = \frac{\cos(\theta_1)}{\cos(\theta_2)}, \qquad \cos(\theta_2) \neq 0$$

(since $\beta_k^{(CD)} = -\dfrac{\|g_{k+1}\|^2}{g_k^T d_k}$)

Then we have three cases:

Case1: If $\cos(\theta_1) = 0$ then

$$\beta_k^{(MLS)} = \beta_k^{(CD)} \tag{14a}$$

Case2: If $\cos(\theta_1), \cos(\theta_2) > 0$ or $\cos(\theta_1), \cos(\theta_2) < 0$ then

$$\beta_k^{(MLS)} = \beta_k^{(CD)} + \frac{\|g_{k+1}\|}{\|d_k\|}\omega^+ \tag{14b}$$

where $\omega^+ = \dfrac{\cos(\theta_1)}{\cos(\theta_2)} > 0$

Case3: If $\cos(\theta_1) < 0$ or $\cos(\theta_2) < 0$ then

$$\beta_k^{(MLS)} = \beta_k^{(CD)} + \frac{\|g_{k+1}\|}{\|d_k\|}\omega^- \tag{14c}$$

where $\omega^- = \dfrac{\cos(\theta_1)}{\cos(\theta_2)} < 0$

3. Outlines of the Proposed Algorithm:

Step(1):*The initial step*: We select starting point $x_0 \in R^n$, and we select the accuracy

solution $\varepsilon > 0$ is a small positive real number and we find $d_k = -g_k$,

$\lambda_0 = Min\,ary(g_0)$, and we set $k = 0$.

Step(2): *The convergence test*: If $\|g_k\| \le \varepsilon$ then stop and set the optimal solution is x_k,

Else, go to step(3).

Step(3): *The line search*: We compute the value of λ_k by Cubic method and that

satisfies the Wolfe conditions in Eqs. (4),(5) and go to step(4).

Step(4): *Update the variables*: $x_{k+1} = x_k + \lambda_k d_k$ and compute $f(x_{k+1}), g_{k+1}$ and

$s_k = x_{k+1} - x_k$, $y_k = g_{k+1} - g_k$.

Step(5): *Check:* if $\|g_{k+1}\| \le \varepsilon$ then stop. Else continue.

Step (6): *The search direction*: We compute the scalar $\beta_k^{(MLS)}$ by using the equations

(14) and (3) and set $k = k+1$, and go to step (4).

4. The Convergence Analysis:

4.1 Sufficient Descent Property:

We will show in this section that the proposed algorithm which is defined in the equations (14) and (3) satisfies the sufficient descent property which satisfies the convergence property.

Theorem (4.1.1):

The search direction d_k that is generated by the proposed algorithm of modified CG satisfies the descent property for all k, when the step size λ_k satisfied the Wolfe conditions (4),(5) .

Proof: we will use the indication to prove the descent property, for $k = 0$, $d_0 = -g_0 \Rightarrow d_0^T g_0 = -\|g_0\| < 0$, then we proved that the theorem is true for $k = 0$, now assume that the theorem is true for any k i.e

$d_k^T g_k < 0$ or $s_k^T g_k < 0$ since $s_k = \lambda_k d_k$, now we will prove that the theorem is true for $k+1$ then

$d_{k+1} = -g_{k+1} + \beta_k^{(MLS)} d_k$

Case1: Multiply both sides of the above equation by g_{k+1} and we set the value of the scalar $\beta_k^{(MLS)}$ in equation (14a) and we get

$$d_{k+1}^T g_{k+1} = -g_{k+1}^T g_{k+1} + (\beta_k^{(MLS)}) g_{k+1}^T d_k \qquad (15)$$

$$d_{k+1}^T g_{k+1} = \|g_{k+1}\|^2 - \frac{\|g_{k+1}\|^2}{g_k^T d_k} g_{k+1}^T d_k$$

$$d_{k+1}^T g_{k+1} = -\|g_{k+1}\|^2 (1 + \frac{g_{k+1}^T d_k}{g_k^T d_k})$$

Then by using the Wolfe condition we get:

$$d_{k+1}^T g_{k+1} \le -\|g_{k+1}\|^2 (1+\rho) \qquad (16)$$

$= -\|g_{k+1}\|^2 (\rho+1)$, $c = (1+\rho) > 0$ sufficient descent satisfied.

Case2: Multiply both sides of the above equation by g_{k+1} and we set the value of the scalar $\beta_k^{(MLS)}$ in equation (14b) and we get

$$d_{k+1}^T g_{k+1} = -g_{k+1}^T g_{k+1} + (\beta_k^{(MLS)}) g_{k+1}^T d_k \qquad (17)$$

$$\beta_k^{(MLS)} = \beta_k^{(CD)} + \frac{\|g_{k+1}\|}{\|d_k\|} \omega^+ \qquad (18)$$

$$\Rightarrow d_{k+1}^T g_{k+1} = -\|g_{k+1}\|^2 - \frac{\|g_{k+1}\|^2 d_k g_{k+1}}{g_k^T d_k} + \frac{\|g_{k+1}\|}{\|d_k\|} g_{k+1}^T d_k \omega^+$$

$$= -\|g_{k+1}\|^2 (1 + \frac{d_k^T g_{k+1}}{g_k^T d_k} - \frac{1}{\|d_k\| \|g_{k+1}\|} g_{k+1}^T d_k \omega^+)$$

Then by using the Wolfe condition we get:

$$\le -\|g_{k+1}\|^2 (1 + \frac{\rho g_k^T d_k}{g_k^T d_k} - \frac{\rho g_k^T d_k}{\|d_k\| \|g_{k+1}\|} \omega^+)$$

$$= -\|g_{k+1}\|^2 (1 + \rho - \frac{\rho g_k^T d_k}{\|d_k\| \|g_{k+1}\|} \omega^+) \qquad (19)$$

Since $g_k^T d_k \le 0$, $c = \left(1 + \rho - \frac{\rho g_k^T d_k}{\|d_k\| \|g_{k+1}\|} \left(\frac{\cos(\theta_1)}{\cos(\theta_2)} \right)^+ \right) > 1$. Then the sufficient descent is satisfied.

Case3: Multiply both sides of the above equation by g_{k+1} and we set the value of the scalar $\beta_k^{(MLS)}$ in equation (14c) and we get

$$d_{k+1}^T g_{k+1} = -g_{k+1}^T g_{k+1} + (\beta_k^{(MLS)}) g_{k+1}^T d_k \tag{20}$$

$$\beta_k^{(MLS)} = \beta_k^{(CD)} + \frac{\|g_{k+1}\|}{\|d_k\|}\omega^- \tag{21}$$

$$\Rightarrow d_{k+1}^T g_{k+1} = -\|g_{k+1}\|^2 - \frac{\|g_{k+1}\|^2 d_k g_{k+1}}{g_k^T d_k} + \frac{\|g_{k+1}\|}{\|d_k\|} g_{k+1}^T d_k \omega^-$$

$$= -\|g_{k+1}\|^2 (1 + \frac{d_k g_{k+1}}{g_k^T d_k} - \frac{1}{\|d_k\|\|g_{k+1}\|} g_{k+1}^T d_k \omega^-) \tag{22}$$

Then by using the Wolfe condition and by using the relation $u^T v = \|u\| \; \|v\| \cos\theta \;\; \Rightarrow u^T v \le \|u\| \; \|v\|$ (where ɵ is the angle between the vectors u and v), we have:

$\le -\|g_{k+1}\|^2 (1 + \frac{\sigma g_k^T d_k}{g_k^T d_k} - \frac{\|d_k\|\|g_{k+1}\|}{\|d_k\|\|g_{k+1}\|}\omega^-)$ sufficient descent condition is satisfied.

Where $c = \left(1 + \rho - \left(\frac{\cos(\theta_1)}{\cos(\theta_2)}\right)^-\right) > 1$

4.3 Global Convergence property:

Assumption:

We assume that:

(i) The level set $S = \{x \in R^n : f(x) \le f(x_0)\}$ is bounded.

(ii) In a neighborhood N of S, the function f is continuously differentiable and its gradient is Lipschitz continuous, i.e. there exists a constant L>0 such that

$$\|g(x) - g(y)\| \le L\|x - y\|, \quad for \quad all \quad x, y \in N.$$

Under these assumptions on f, there exists a constant $\Gamma \ge 0$ such that $\|g(x)\| \le \Gamma$,

In [1,3] it is proved that, for any conjugate gradient method with strong Wolfe line search, the following general result holds.

Lemma 4.3.1:

Let assumptions (i) and (ii) hold and consider any conjugate gradient method (2) and (3), where d_k is a descent direction and λ_k is obtained by the strong Wolfe line search. If

$$\sum_{k\geq 1}\frac{1}{\|d_k\|^2}=\infty \qquad (23)$$

Then

$$\lim_{k\to\infty}\inf\|g_k\|=0 \qquad (24)$$

For uniformly convex functions which satisfy the above assumptions, we can prove that the norm of d_{k+1} given by (3) is bounded above. Assume that the function f is a uniformly convex function.

Using lemma 4.3.1 the following result can be proved.

Theorem 4.3.2:

Suppose that the assumptions (i) and (ii) hold. Consider the algorithm (2), (15). If $\|d_k\|$ tends to zero and there exists nonnegative constants $\eta 1$ and $\eta 2$ such that:

$$\|g_k\|^2 \geq \eta 1\|d_k\|^2, \qquad \|g_{k+1}\|^2 \geq \eta 2\|d_k\| \qquad (25)$$

and *f* is a uniformly convex function, then

$$\lim_{k\to\infty}\inf\|g_k\|=0 \qquad (26)$$

Proof:

Case1: In this case we have:

$$\beta_k^{(MLS)}=\beta_k^{(CD)}$$

From eq.(25), we get:

$$\left|\beta_k^{MLS}\right| \leq \left|-\frac{\|g_{k+1}\|^2}{g_k^T d_k}\right| \leq \frac{\eta 2\|d_k\|}{\eta 1\|d_k\|^2}$$

since

$$\|d_{k+1}\| \leq \|g_{k+1}\| + \left|\beta_k^{MLS}\right|\|d_k\|$$

$$\|d_{k+1}\| \leq \Gamma + \frac{\eta 2}{\eta 1}$$

$$\sum_{k\geq 1}\frac{1}{\|d_{k+1}\|^2}=\infty$$

$$\frac{1}{\left(\Gamma+\frac{\eta 2}{\eta 1}\right)^2}\sum_{k\geq 1}1=\infty$$

Then:

$$\liminf_{k\to\infty}\|g_{k+1}\|=0$$

Case 2: In this case we have:

$$\beta_k^{MLS}=-\frac{\|g_{k+1}\|^2}{g_k^T d_k}+\frac{\|g_{k+1}\|}{\|d_k\|}w^+$$

$$\left|\beta_k^{MLS}\right|\le\left|-\frac{\|g_{k+1}\|^2}{g_k^T d_k}+\frac{\|g_{k+1}\|}{\|d_k\|}w^+\right|$$

From eq.(25), we get:

$$\left|\beta_k^{MLS}\right|\le\frac{\eta 2\|d_k\|}{\eta 1\|d_k\|^2}+\frac{\Gamma w^+}{\|d_k\|}=\frac{\eta 2}{\eta 1\|d_k\|}+\frac{\Gamma w^+}{\|d_k\|}$$

since

$$\|d_{k+1}\|\le\|g_{k+1}\|+\left|\beta_k^{MLS}\right|\|d_k\|$$

$$\|d_{k+1}\|\le\Gamma+\left(\frac{\eta 2}{\eta 1\|d_k\|}+\frac{\Gamma w^+}{\|d_k\|}\right)\|d_k\|=\Gamma+\left(\frac{\eta 2}{\eta 1}+\Gamma w^+\right)$$

$$\sum_{k\ge 1}\frac{1}{\|d_{k+1}\|^2}=\infty$$

$$\frac{1}{\left(\Gamma+\left(\frac{\eta 2}{\eta 1}+\Gamma w^+\right)\right)^2}\sum_{k\ge 1}1=\infty$$

Then:

$$\liminf_{k\to\infty}\|g_{k+1}\|=0$$

Case 3:

$$\beta_k^{MLS}=-\frac{\|g_{k+1}\|^2}{g_k^T d_k}+\frac{\|g_{k+1}\|}{\|d_k\|}w^-$$

$$\left|\beta_k^{MLS}\right|\le\left|-\frac{\|g_{k+1}\|^2}{g_k^T d_k}+\frac{\|g_{k+1}\|}{\|d_k\|}w^-\right|$$

By the similar way and by use the absolute value:

$$\sum_{k\ge 1}\frac{1}{\|d_{k+1}\|^2}=\infty$$

$$\frac{1}{\left(\Gamma+\left(\frac{\eta 2}{\eta 1}+\Gamma w^-\right)\right)^2}\sum_{k\ge 1}1=\infty$$

Then:

$$\liminf_{k\to\infty}\|g_{k+1}\|=0$$

5. Computational and Results:

In this section, we reported some numerical results obtained with the implementation of the new algorithm on a set of unconstrained optimization test problems. We have selected (10) large scale unconstrained optimization problems in extended or generalized form, for each test function we have considered numerical experiment with the number of variable n=1000,5000,10000.

Using the standard Wolfe line search conditions (4), (5) , the stopping criteria is the $\|g_{k+1}\| \le 10^{-6}$. The programs were written in FORTRAN 90.

We compare our method namely (MLS) with the FR method (7). The preliminary numerical results of our tests are reported in Tables (1),(2) and (3). The first column "test fun." The names of test functions, the second column "NOI" denoted the number of iterations, the third column "NOF" denoted the number of calculated functions and the fourth column "MIN" denoted the minimum value. We compute:

$$\cos\theta 1 = \frac{-g_{k+1}^T g_k}{\|g_{k+1}\|\|g_k\|} \quad \text{and} \quad \cos\theta 2 = \frac{-g_k^T d_k}{\|g_k\|\|d_k\|}.$$

Table (1)

Comparative Performance of the Two Algorithms for Group of Test Functions at N=1000

Test Fun.	**FR-CG algorithm**			**MCG algorithm**		
	NOI	NOF	MIN	NOI	NOF	MIN
Non-diagonal	148	345	1.53432E-013	24	59	2.770565E-027
Wolfe	192	385	5.265267E-014	83	167	2.736934E-014
Wood	3354	17377	1.102234E-013	76	158	4.556063E-015
Rosen	175	440	8.234663E-014	27	71	6.312248E-018
Cubic	62	132	1.387285E-013	13	35	2.501331E-017
Edgar	6	14	1.0135803E-014	7	18	1.049656E-017
Sum	31	173	7.6506322E-009	24	136	4.897356E-009
Powell3	1002	2009	7.648563E-009	164	350	4.351358E-010
Dixon	396	795	9.809334E-014	35	73	3.992565E-014
Reciep	11	30	5.675117E-015	13	59	8.318482E-013
Total	5377	21700		466	1126	

Table (2)

Comparative Performance of the Two Algorithms for Group of Test Functions at N=5000

Test Fun.	**FR-CG algorithm**			**MCG algorithm**		
	NOI	NOF	MIN	NOI	NOF	MIN
Non-diagonal	109	266	5.84975E-015	24	59	9.858789E-028
Wolfe	208	419	2.584547	117	236	2.584547
Wood	4887	27294	2.20567E-014	79	164	2.454351E-015
Rosen	182	454	4.89090E-014	27	71	3.156147E-017
Cubic	63	134	6.93352E-013	13	35	1.250665E-016
Edgar	6	14	5.06790E-014	7	18	5.248294E-017
Sum	44	252	8.28405E-009	34	202	5.769684E-009
Powell3	4915	9835	1.69918E-010	84	190	3.522151E-010
Dixon	1999	4000	0.5000000	35	73	8.009199E-014
Reciep	11	30	2.83926E-014	14	37	3.162427E-013
Total	12424	42698		434	1085	

Table (3)

Comparative Performance of the Two Algorithms for Group of Test Functions at N=10000

Test Fun.	FR-CG algorithm			MCG algorithm		
	NOI	NOF	MIN	NOI	NOF	MIN
Non-diagonal	138	325	5.298103E-014	24	60	3.00723E-026
Wolfe	239	481	2.584548	134	272	2.584549
Wood	7969	34587	1.821634E-014	80	166	4.038886E-015
Rosen	183	456	9.625753E-014	27	71	6.312159E-017
Cubic	63	134	1.386707E-012	13	35	2.5013139E-016
Edgar	6	14	1.013580E-013	7	18	1.049658E-016
Sum	54	258	1.368982E-008	37	175	7.598072E-009
Powell3	5648	11301	2.099293E-010	239	504	9.6432081E-010
Dixon	528	1058	0.5000000	34	71	4.4836988E-014
Reciep	11	30	5.680229E-014	16	44	7.684872E-014
Total	14839	48644		611	1416	

6. References:

[1] Andrei, N., (2007), "Scaled conjugate gradient algorithms for unconstrained optimization" Comput. Optim. Appl. 38, pp.401-416.

[2] Dai, Y. H. and Yuan, Y., (1999), " A nonlinear conjugate gradient method with a strong global convergence property", SIAM J. Optim., 10 , pp. 177-182.

[3] Dai, Y. H. and Liu, G.H., Sun, D.F., Yin, X., and Yuan, Y., (1999), "Convergence properties of nonlinear conjugate gradient methods. SIAM J. Optim. 10, pp 348-358.

[4] Fletcher, R. and Reeves, C., (1964), " Function minimization by conjugate gradients", Comput. J., 7, pp. 149-154.

[5] Fletcher, R., (1987), " Practical Methods of Optimization" vol. 1: Unconstrained Optimization, John Wiley & Sons, New York.

[6] Hestenes, M. R. ,Stiefel, E.L. (1952), "Methods of Conjugate gradient for solving linear systems",J. Res. Ntal.Bur.stant., 49,pp. 409-346.

[7] Kiwiel, K.C., Murty, K., (1996) "Convergence of the steepest descent method for minimizing quasiconvex functions. J. Optim. Theory Appl. 89 (1), pp.221-226.

[8] Polak, E. and Ribiere, G., (1969), " Note sur la convergence de directions conjugees", Rev. Francaise Informat Recherche Opertionelle, 3e Annee 16, pp. 35-43.

[9] Polyak, B. T.,(1969), "The conjugate gradient method in extreme problems", USSR Comp Math. Phys., 9, pp. 94-112.

[10] Liu, Y. and Storey, C., (1991), "Efficient generalized conjugate gradient algorithms", Part 1: Theory, J. Optim. Theory Appl., 69, pp. 129-137.